Machine Learning - NextStep

Como podemos ver durante o segundo filme da saga “De Volta para o Futuro”, o personagem Biff encontra o *Grays Sports Almanac*, contendo resultados de todos os eventos esportivos do ano de 1950 até 2000 em uma lixeira no ano de 2015. Inflado pelo seu egoísmo e ganância, ele leva para o seu “eu” do passado no ano de 1955. Com o passar do tempo, Biff constrói toda a sua fortuna apostando nos mais variados esportes e é ajudado pelo fato de sempre saber o resultado de determinado evento.

Quem não queria estar na mesma situação do Biff? Sabemos que viagem no tempo é algo além do nosso alcance e que não existe um livro como esse. Porém, com os diversos avanços da inteligência artificial, especialmente no que diz respeito à Machine Learning, somos capazes de criar algoritmos que conseguem prever possíveis resultados através do treinamento de redes neurais com a ajuda de dados históricos.

Como o almanaque do livro possui resultados de diversos eventos esportivos, propusemos um algoritmo de aprendizagem de máquina, com base na Regressão Logística, que tem como fundamento prever os resultados de partidas do Campeonato Brasileiro da Série A, como uma tentativa de replicar o livro usado por Biff.

Para treinamento, optamos por uma base de dados que contém as temporadas dos anos de 2012 até 2017, apresentando informações essenciais para a Regressão Logística, tais como: Time Mandante, Time Visitante, quantos gols cada um marcou e uma tabela de Resultados.

A coluna chave desta base de dados é, certamente, a de Resultados, a qual possui a seguinte estrutura:

0: Empate

1: Vitória do Time Mandante

2: Vitória do Time Visitante

Pudemos perceber que, além dos resultados desta coluna, o arquivo continha diversos cálculos que são fundamentais no treinamento, como: Média de Gols do Time Mandante, Média de Gols do Time Visitante, Probabilidade de Vitórias do Mandante, do Visitante, bem como do Empate.

Sendo assim, para a comprovação da eficácia deste modelo, tentaremos prever os resultados de uma parte da própria base de dados e visualizá-los, justapostos com os dados reais, através de uma matriz de confusão. Finalmente, saberemos de imediato a eficácia da nossa previsão.

Link para a Base de Dados : <https://docs.google.com/spreadsheets/d/18MjS6gasfVdedCywqMiS9DrAdHtJtXCq0mA9Mzt-qN4/edit?usp=sharing>